Exact marginals and normalizing constant for Gibbs distributions

Cécile Hardouin, Xavier Guyon

Exact marginals and normalizing constant for Gibbs distributions

C. Hardouin [a,1], X. Guyon [a]
[a] *CES/SAMOS-MATISSE/Université de Paris 1*

Abstract

We present a recursive algorithm for the calculation of the marginal of a Gibbs distribution π. A direct consequence is the calculation of the normalizing constant of π.

Résumé

Récurrences et constante de normalisation pour des modèles de Gibbs. Nous proposons dans ce travail une récurrence sur les lois marginales d'une distribution de Gibbs π. Une conséquence directe est le calcul exact de la constante de normalisation de π.

1. Introduction

Usually, obtaining the marginals and/or the normalizing constant C of a discrete probability distribution π involves high dimensional summation : for example, for the binary Ising model on a simple grid 10×10, the calculation of C involves 2^{100} terms. One way to prevent this problem is to change distribution of interest for an alternative as, for example in spatial statistics, replacing the likelihood for the conditional pseudo likelihood ([1]). Another solution consists of estimating the normalizing constant; see for example Pettitt & al ([8]) and Moeller & al ([7]) for efficient Monte Carlo methods, Bartolucci and Besag ([2]) for a recursive algorithm computing the exact likelihood of a Markov random field, Reeves and Pettitt ([9]) for an efficient computation of the normalizing constant for a factorisable model.

We present specific results for a Gibbs distribution π. We derive results of Khaled ([5,6]) who gives an original linear recursion on the marginals of π, the law of $Z = (Z_1, Z_2, \cdots, Z_T) \in E^T$; this result eases the calculation of π's normalizing constant. We generalize Khaled results noticing that if π is a Gibbs

Email addresses: Cecile.Hardouin@univ-paris1.fr (C. Hardouin), guyon@univ-paris1.fr (X. Guyon).
[1] Adresse de correspondance : C. Hardouin, SAMOS-MATISSE/Université de Paris 1, 90 rue de Tolbiac, 75634 Paris Cedex 13

distribution on $\mathcal{T} = \{1, 2, \cdots, T\}$, then π is a Markov field on $\mathcal{T}$, so it is easy to manipulate its conditional distributions that are the basic tools of our forward recursions.

2. Markov representations of a Gibbs field

Let $T > 0$ be a fix positive integer, $E = \{e_1, e_2, \cdots, e_N\}$ a finite state space, $Z = (Z_1, Z_2, \cdots, Z_T) \in E^T$ a temporal sequence with distribution π. Let us denote $z(t) = (z_1, z_2, \cdots, z_t)$. We assume that π is a Gibbs distribution with energy and potentials:

$$\pi(z(T)) = C \exp U_T(z(T)) \text{ with } C^{-1} = \sum_{z(T)\in E^T} \exp U_T(z(T)) \text{ where} \tag{1}$$

$$U_t(z(t)) = \sum_{s=1,t} \theta_s(z_s) + \sum_{s=2,t} \Psi_s(z_{s-1}, z_s) \text{ for } 2 \leq t \leq T, \text{ and } U_1(z_1) = \theta_1(z_1).$$

So, π is a bilateral 2 nearest neighbours Markov field ([4,3])

$$\pi(z_t \mid z_s, 1 \leq s \leq T \text{ and } s \neq t) = \pi(z_t \mid z_{t-1}, z_{t+1}) \tag{2}$$

but Z is also a Markov chain :

$$\pi(z_t \mid z_s, s \leq t-1) = \pi(z_t \mid z_{t-1}) \text{ if } 1 < t \leq T. \tag{3}$$

An important difference appears between formulas (3) and (2): indeed, (2) is computationnally feasible, when (3) is not.

3. Recursion over marginal distributions

3.1. *Future-conditional contribution* $\Gamma_t(z(t))$

For $t \leq T-1$, the distribution $\pi(z_1, z_2, \cdots, z_t \mid z_{t+1}, z_{t+2}, \cdots, z_T)$ conditionally to the future, depends only on z_{t+1}:

$$\pi(z_1, z_2, \cdots, z_t \mid z_{t+1}, z_{t+2}, \cdots, z_T) = \frac{\pi(z_1, z_2, \cdots, z_T)}{\sum_{u_1^t \in E^t} \pi(u_1^t, z_{t+1}, ..z_T)} = \pi(z_1, z_2, \cdots, z_t \mid z_{t+1}).$$

We can also write $\pi(z_1, z_2, \cdots, z_t \mid z_{t+1}) = C_t(z_{t+1}) \exp U_t^*(z_1, z_2, \cdots, z_t; z_{t+1})$ where U_t^* is the future-conditional energ :

$$U_t^*(z_1, z_2, \cdots, z_t; z_{t+1}) = U_t(z_1, z_2, \cdots, z_t) + \Psi_{t+1}(z_t, z_{t+1}), \tag{4}$$

and $C_{t+1}(z_{t+1})^{-1} = \sum_{u_1^t \in E^t} \exp\{U_t^*(u_1, ..., u_t; z_{t+1})\}$. Then, for $i = 1, N$:

$$\pi(z_1, z_2, \cdots, z_t \mid z_{t+1} = e_i) = C_t(e_i)\gamma_t(z_1, z_2, \cdots, z_t; e_i) \text{ where } \gamma_t(z(t); e_i) = \exp U_t^*(z(t); e_i).$$

With the convention $\Psi_{T+1} \equiv 0$, we define for $t \leq T$, the vector $\Gamma_t(z(t)) \in \mathbb{R}^N$ of the *future-conditional contributions* as

$$(\Gamma_t(z(t)))_i = \gamma_t(z(t); e_i),\ 1 \leq i \leq N.$$

and the recursion matrix A_t by

$$A_t(i,j) = \exp\{\theta_t(e_j) + \Psi_{t+1}(e_j, e_i)\},\ i, j = 1, N. \tag{5}$$

Then we get the following fundamental recurrence.

Proposition 3.1 *For all* $2 \leq t \leq T$, $z(t) = (z_1, z_2, \cdots, z_t) \in E^t$ *and* $e_i \in E$, *we have:*

$$\gamma_t(z(t-1), e_j; e_i) = A_t(i,j) \times \gamma_{t-1}(z(t-1); e_j) \ , \tag{6}$$

and

$$\sum_{z_t \in E} \Gamma_t(z(t-1), z_t) = A_t \Gamma_{t-1}(z(t-1)). \tag{7}$$

3.2. *Forward recursions on marginals and normalization constant*

Let us define the following $1 \times N$ row vectors : $E_1 = B_T = (1, 0, \cdots, 0)$, and the $(B_t)_{t=T,2}$ defined by the forward recursion $B_{t-1} = B_t A_t$ if $t \leq T$; we also denote $K_1 = \sum_{z_1 \in E} \Gamma_1(z_1) \in \mathbb{R}^N$. We give below the main result of this work.

Proposition 3.2 *Marginal distributions* π_t *and calculation of the normalization constant* C.
(1) For $1 \leq t \leq T$ *:*

$$\pi_t(z(t)) = C \times B_t \Gamma_t(z(t)). \tag{8}$$

(2) The normalization constant C *of the joint distribution* π *verifies:*

$$C^{-1} = E_1 A_T A_{T-1} \cdots A_2 K_1. \tag{9}$$

The formula (9) reduces to $C^{-1} = E_1 A_T A^{T-2} K_1$ for time invariant potentials.

As a basic example, let us consider $E = \{0, 1\}$, $\theta_t(z_t) = \alpha z_t$, and $\Psi_{t+1}(z_t, z_{t+1}) = \beta z_t z_{t+1}$; the analytic expressions of A, K_1 are trivially derived. We computed $C^{-1} = E_1 A_T A^{T-2} K_1$ for increasing values of T; the computing time is always negligible for $T \leq 700$, whereas computing C^{-1} by direct summation needs 750 seconds for $T = 20$, 6 hours for $T = 25$, and the method becoming ineffectual for $T > 25$.

4. Extensions to general Gibbs fields

There are various generalisations of the preceeding results.

4.1. *Temporal Gibbs model*

Let us give the following example as an illustration to possible extensions. Coming back to the previous model (1), we add the interaction potentials $\Psi_{2,s}(z_{s-2}, z_s)$. Then π is a 4 nearest neighbours Markov field but also a Markov chain of order 2. Conditionally to the future, we get

$$\pi(z_1, z_2, \cdots, z_t \mid z_{t+1}, z_{t+2}, \cdots, z_T) = \pi(z(t) \mid z_{t+1}, z_{t+2}) = C_t(z_{t+1}, z_{t+2}) \exp U_t^*(z(t); z_{t+1}, z_{t+2}), \text{ with}$$
$$U_t^*(z(t); z_{t+1}, z_{t+2}) = U_t(z(t)) + \Psi_{1,t+1}(z_t, z_{t+1}) + \Psi_{2,t+1}(z_{t-1}, z_{t+1}) + \Psi_{2,t+2}(z_t, z_{t+2}),$$

Then, for a, b and $c \in E$, $U_t^*(z(t-1), a; (b,c)) = U_{t-1}^*(z(t-1); (a,b)) + \theta_t(a) + \Psi_{1,t+1}(a,b) + \Psi_{2,t+2}(a,c)$; analogously to the previous example, we define the future-conditional contributions and the $N^2 \times N^2$ matrices A_t by

$$\gamma_t(z(t); (z_{t+1}, z_{t+2})) = \exp U_t^*(z(t); (z_{t+1}, z_{t+2})$$
$$A_t((i,j),(k,i)) = \exp\{\theta_t(e_k) + \Psi_{1,t+1}(e_k, e_i) + \Psi_{2,t+2}(e_k, e_j)\}$$

Similarly as 3.1, we get the following recursion:

$$\gamma_t(z(t-1), e_k; (e_i, e_j)) = A_t((i,j),(k,i)) \times \gamma_{t-1}(z(t-1); (e_k, e_i))$$

We thus obtain a recurrence (7) on the contributions $\Gamma_t(z(t))$ and analogous results as (8) and (9) for the bivariate Markov chain (Z_{t-1}, Z_t), $t = 1, T$.

4.2. *Spatial Gibbs fields*

For $t \in \mathcal{T} = \{1, 2, ...T\}$, let us consider $Z_t = (Z_{(t,i)},\ i \in \mathcal{I})$, where $\mathcal{I} = \{1, 2, \cdots, m\}$, $Z_{(t,i)} \in F$. Then $Z = (Z_s,\ s = (t,i) \in \mathcal{S})$ is a spatial field on $\mathcal{S} = \mathcal{T} \times \mathcal{I}$. We note again $z_t = (z_{(t,i)},\ i \in \mathcal{I})$, $z(t) = (z_1, .., z_t)$, $z = z(T)$ and we suppose that the distribution π of Z is a Gibbs distribution with translation invariant potentials $\Phi_{A_k}(\bullet),\ k = 1, K$ associated to a family of subsets $\{A_k, k = 1, K\}$ of $\mathcal{S}$. For $A \subseteq \mathcal{S}$, let us define $H(A) = \sup\{|u - v|, \exists (u,i) \text{ and } (v,j) \in A\}$, and $H = \sup\{H(A_k),\ k = 1, K\}$. With this notation, we write the Gibbs-energy

$$U(z) = \sum_{h=0}^{H} \sum_{t=h+1}^{T} \Psi(z_{t-h}, \cdots, z_t) \text{ with } \Psi(z_{t-h}, \cdots, z_t) = \sum_{k:H(A_k)=h} \ \sum_{s \in S_t(k)} \Phi_{A_k+s}(z)$$

where $S_t(k) = \{s = (u,i) : A_k + s \subseteq \mathcal{S} \text{ and } t - H(A_k) \leq u \leq t\}$. Then (Z_t) is a Markov process of order H and $Y_t = (Z_{t-H}, Z_{t-H+1}, \cdots Z_t)$, $t > H$ a Markov chain on E^H for which we get the results (8) and (9).

We applied the result to the calculation of the normalization constant for an Ising model. For $m = 10$ and $T = 100$, the computing time is less than 20 seconds.

References

[1] J. Besag, 1974. Spatial interactions and the statistical analysis of lattice systems. *J. Roy. Statist. Soc.* **B** 148, 1-36

[2] F. Bartolucci, J. Besag, 2002. A recursive algorithm for Markov random fields. *Biometrika* **89** 3, 724-730

[3] X. Guyon, 1995. *Random Fields on a Network: Modeling, Statistics, and Applications.* Springer-Verlag, New York

[4] R. Kindermann, J.L. Snell, 1980.), *Markov random fields and their applications*, Contemp. Maths.

[5] M. Khaled, 2008. A multivariate generalization of a Markov switching model. *working paper C.E.S. Université Paris 1, http://mkhaled.chez-alice.fr/khaled.html*

[6] M. Khaled, 2008. Estimation bayésienne de modèles espace-état non linéaires. *Ph. D. Thesis, Université Paris 1, http://mkhaled.chez-alice.fr/khaled.html*

[7] J. Moeller, A.N. Pettitt, R. Reeves, K.K. Berthelsen, 2006. An efficient Markov chain Monte Carlo method for distributions with intractable normalizing constants. *Biometrika* **93** 2 , 451-458.

[8] A.N. Pettitt, N. Friel, R. Reeves, 2003. fficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice. *J. Roy. Statist. Soc.* **B** 65, part 1, 235-246

[9] R. Reeves,A.N. Pettitt, 2004. Efficient recursions for general factorisable models. *Biometrika* **91** 3 , 751-757.